Gaussian process tomography for soft-X ray spectroscopy at WEST without equilibrium information

T. Wang[1,2,3,a)], D. Mazon[2], J. Svensson[4], D. Li[1], A. Jardin[2] and G. Verdoolaege[3,5]

[1]*Southwestern Institute for Physics, CNNC, C-610200 Chengdu, China*
[2]*Institute for Magnetic Fusion Research, CEA, F-13115 Saint-Paul-lez-Durance, France*
[3]*Department of Applied Physics, Ghent University, B-9000 Ghent, Belgium*
[4]*Max-Planck-Institut für Plasmaphysik, D-17491 Greifswald, Germany*
[5]*Laboratory for Plasma Physics, Royal Military Academy (LPP-ERM/KMS), B-1000 Brussels, Belgium*

(Received XXXXX; accepted XXXXX; published online XXXXX)
(Dates appearing here are provided by the Editorial Office)

Gaussian process tomography (GPT) is a recently developed tomography method based on Bayesian probability theory.[1, 2] By modeling the SXR emissivity field in a poloidal cross-section as a Gaussian process, Bayesian SXR tomography can be carried out in a robust and extremely fast way. Owing to the short execution time of the algorithm, GPT is an important candidate for providing real-time reconstructions with a view to impurity transport and fast MHD control. In addition, the Bayesian formalism allows quantifying uncertainty on the inferred parameters. In this paper, the GPT technique is validated using a synthetic data set expected from the WEST tokamak and results are shown of its application to the reconstruction of SXR emissivity profiles measured on Tore Supra. The method is compared with the standard algorithm based on minimization of the Fisher information.

I. Introduction

Nuclear fusion research, aiming at the development of a sustainable and safe energy source, is presently largely concentrated around the construction and operation of the ITER magnetic confinement device. One of the primary issues threatening safe and efficient operation of this type of machines, called tokamaks, is the accumulation of impurities in the plasma core, causing fuel dilution and radiative power loss, potentially leading to a complete loss of plasma confinement in a disruption. An important source of impurity originates from the interaction of the hot hydrogenic plasma with the wall components. In ITER, tungsten ions (W) may pose a risk because highly charged impurities radiate energy very efficiently. In reactor-relevant plasmas, thermonuclear burn will only tolerate tungsten concentrations less than 10^{-4}, in order to avoid plasma disruptions.[3] Therefore, a detailed understanding of core impurity transport in tokamaks, including the interplay with magnetohydrodynamic (MHD) activity, is crucial. This requires reliable information about impurity distributions, at a time resolution that is adapted to MHD time scales.

a) Electronic mail: tianbo.wang@ugent.be.

Soft X-ray (SXR) spectroscopy is a diagnostic technique that has the potential to deliver valuable information in this respect.[4,5] This diagnostic can provide very good temporal resolution (up to 1 MHz), which is sufficient for MHD activity and impurity transport studies. Particularly, the plasma is optically thin for soft X-ray radiation in the range from 1 keV to 15 keV, which makes SXR tomography a powerful tool for studying core plasma physics.[6] The technique was pioneered by groups in several laboratories using a single camera by means of plasma rotation.[7,8,9,10] Full tomography with at least two cameras was realized soon afterwards,[11,12] followed by systems with even more cameras.[13,14,15,16]

In the past, various tomographic reconstruction techniques have been applied to SXR, such as the Cormack method,[17] the maximum entropy method,[18] the minimum Fisher information method,[19] etc. Particularly the minimum Fisher information technique has been widely adopted in the fusion community. This reconstruction method involves χ^2 optimization, regularized by the Fisher information. Intuitively, the goal is to find the least complex solution that is compatible with the data. The method is often implemented, e.g. on Tore Supra and WEST,[20] using additional information concerning the location of the equilibrium magnetic flux surfaces, obtained from magnetic measurements. It provides a good trade-off between the initial magnetic flux surface information and the SXR data.

However, the assumption about the equilibrium magnetic flux surfaces may be too restrictive. For instance, in the presence of poloidally asymmetric impurity distributions, it would be better to rely solely on the SXR measurements for a faithful reconstruction of impurity concentration profiles. Furthermore, the equilibrium information might be inaccurate during various events that can partially break the equilibrium condition, especially in the plasma core area, where uncertainty levels are relatively large. On the other hand, the equilibrium information can be useful as it acts as a soft constraint (prior information in the Bayesian sense), providing further regularization of the inherently ill-posed tomographic problem on top of other regularization schemes.

In this paper we introduce a recently developed algorithm using Bayesian inference[1]. Earlier, this method was applied successfully to a soft X-ray (SXR) diagnostic in the stellarator Wendelstein 7-AS,[2] and it has also been applied to current tomography, bolometry and interferometry at JET and W7-X. The main novelty and attraction lies in the effective regularization technique by a Gaussian process prior, enabling SXR tomographic reconstructions without any additional information on the magnetic equilibrium. A further important advantage is its computational efficiency, rendering the method sufficiently fast for real-time applications. In Section II, first a newly developed SXR diagnostic system using gas electron multipliers (GEM) is presented. The mathematical formalism of Bayesian inference and Gaussian process tomography is treated in Section III. Next, validation of the method using a synthetic phantom is discussed in Section IV, and a comparison is shown of the present Bayesian tomography method with the minimum Fisher information method. The conclusion is presented in Section V.

II. WEST soft X-ray diagnostic

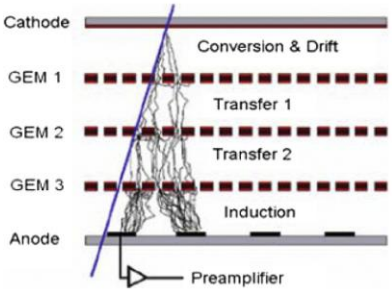

Fig. 1. Schematic of the triple GEM.

WEST – Tungsten (W) Environment in Steady-State Tokamak – is a recently upgraded medium-size tokamak equipped with an actively cooled tungsten divertor. The unique features of WEST lie in its long-pulse capabilities and operation with actively cooled components and superconducting magnets inherited from Tore Supra, combined with a high level of auxiliary power. The WEST tungsten divertor elements, which are the key components facing the largest part of the heat and particle fluxes coming from the plasma during experiments, have the same design and manufacturers as the equivalent ITER components.

The WEST SXR diagnostic system is presently being commissioned with two triple-gas electron multiplier (GEM) cameras, located in the same poloidal cross-section to allow tomographic reconstruction.[21] The triple-GEM detector is based on photoionization in a flowing gas mixture subjected to an electric field and enclosed in Mylar foil. As depicted in Fig. 1, photoelectrons are produced in the first conversion layer and drift towards a perforated copper-clad Kapton foil. A voltage is applied to the foil, causing electron avalanching, hence amplifying the detector signal. The process is repeated in two successive GEM foils, followed by charge collection on the anode strips (pixels). The GEM detectors work in photon counting mode with energy discrimination. Compared to photodiode detectors, the GEM concept separates the regions where photoionization, amplification and detection takes place. An additional advantage is that electrons travel fast to the anode in about 50 ns, while the GEM holes are ion-free after ca. 1 µs. Therefore, the system has high-rate capabilities of ca. 10^6-10^7 ph.s^{-1}.mm^{-2}. Further advantages of GEM detectors are their compactness, good spatial and temporal resolution and good neutron-resistance. As such, the GEM detection system is a good candidate for SXR measurement in ITER and future reactors. As shown in Fig. 2, one of the cameras of the WEST SXR diagnostic views along the horizontal direction through 128 lines-of-sight (LOS) from the low-field-side to the high-field-side. The other camera is located at the top of the device, viewing downwards along 75 lines-of-sight. Hence, the majority of the core plasma region is covered with a good spatial resolution ($\sim$1 cm in the equatorial plane). The GEM system can provide a temporal resolution for real-time analysis of 1 kHz (five energy windows within an energy range of 2-15 keV), while the full spectrum will be stored off-line at a rate of up to 10 kHz, with a view to more detailed analysis and study of fast plasma phenomena. This setup provides good capabilities for studying fast MHD activity and impurity transport, in particular for tungsten transport.

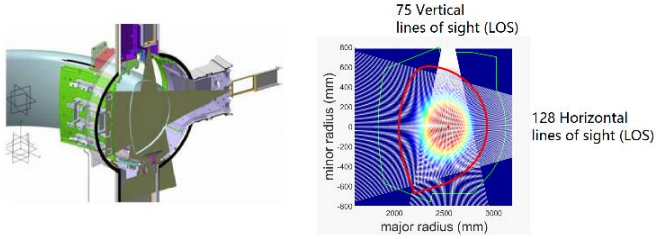

Fig. 2. Tomographic capabilities of the WEST SXR system based on GEM detectors. The horizontal camera views along 128 lines-of-sight. The vertical camera is inside the vertical port and is coupled to 75 sight lines.

The purpose of SXR tomography in magnetic fusion devices is to reveal the spatial distribution of SXR emissivity in a poloidal cross-section, by inversion of a number of noisy line-integrated emissivity measurements. For a Maxwellian plasma and with a spectrum dominated

by hydrogenic bremsstrahlung, the SXR radiation power density $d\varepsilon$ per photon energy interval dE is given by

$$\frac{d\varepsilon}{dE_{ff}} \propto Z_{eff}\, {n_e}^2\, T_e^{\frac{1}{2}}\, e^{-\frac{E}{T_e}}\, g_{ff}(T_e, E)\ . \quad (1)$$

Here, T_e is the electron temperature, n_e is the electron density, Z_{eff} is the effective charge number and g_{ff} is the Gaunt factor which is a function of T_e and E. A common and simple approach to discretize the emissivity field in a poloidal cross-section uses a square grid. We here impose a 100 × 100 grid comprised of square cells with a dimension of 16 mm × 16 mm. The SXR emissivity within each pixel can reasonably be assumed to be constant, so the SXR line-integrated emissivities $\bar{d}_m$ along m viewing chords can be written in the following matrix form:

$$\bar{d}_m = \bar{\bar{R}}_{m\times n} \cdot \bar{E}_n + \bar{\varepsilon}. \quad (2)$$

Here, $\bar{E}_n$ is the unknown vector of local emissivities in $n = 10^4$ cells, while $\bar{\bar{R}}$ is the geometry matrix, whose elements R_{ij} represent the physical length of chord i through cell j. $\bar{\varepsilon}$ denotes an error term to account for measurement uncertainty, which is usually limited to statistical errors only.

III. Probabilistic model and parameter estimation

The tomography problem essentially involves the prediction of high-dimensional physics parameters by inversion of a limited number of measurements. This is an ill-posed problem, as the number of measurements (SXR line integrals) is always lower than the number of unknowns (emissivity value in each cell). There exists a variety of reconstruction algorithms to solve the inversion problem, some of which have already been mentioned in the introduction. They can be divided in two main categories: optimization methods and probabilistic methods. Given a forward model like the one in Eq. (2), the optimization criterion minimizes the difference between the measured line integrals and the prediction by the model. Because of the ill-posedness, the optimization has to be combined with some regularization technique, e.g. assuming a spline model for the local emissivity field, or by optimizing at the same time some information measure like the Shannon entropy or the Fisher information.

a. Bayesian inference

In this paper we choose the probabilistic methodology, which provides a probability distribution $p(\bar{E}_n)$ of the emissivity in all cells rather than a single solution. In Bayesian inference, one starts from the prior probability distribution of the emissivity field, which can be used to encode the regularization (see Section III.b). This is then updated through Bayes' theorem as data become available:

$$p(\bar{E}_n|\bar{d}_m) = \frac{p(\bar{d}_m|\bar{E}_n)\, p(\bar{E}_n)}{p(\bar{d}_m)} \sim p(\bar{d}_m|\bar{E}_n)\, p(\bar{E}_n), \quad (3)$$

$$p(\bar{d}_m) = \int p(\bar{d}_m, \bar{E}_n) d\bar{E}_n = \int p(\bar{d}_m|\bar{E}_n)\, p(\bar{E}_n)\, d\bar{E}_n \quad (4)$$

$\bar{E}_n$ *Vector of emissivity values in all n pixels at a particular time t*

$\bar{d}_m$ *Vector of m line-integrated GEM array measurements at time t*

In Eq. (3), the likelihood term $p(\bar{d}_m|\bar{E}_n)$ measures the mismatch between the measured line integrals $\bar{d}_m$ and their predictions (2) by the forward model, under the assumption of some emissivity field $\bar{E}_n$. The evidence (marginal likelihood) $p(\bar{d}_m)$ depends on the particular forward measurement model, which we will assume to be fixed. Therefore it can be considered as a normalization factor, independent of the emissivity. The posterior probability distribution $p(\bar{E}_n|\bar{d}_m)$ quantifies our uncertainty on the estimated emissivity field, given our model, prior knowledge and the measured data. Thus, Bayesian inference yields probabilities for all possible results consistent with our model. In principle, systematic uncertainties can also be estimated, provided some knowledge is available about them from other sources of information, such as other experiments. Another important advantage is the ease with which heterogeneous sources of information can be integrated into a single coherent model. This is particularly relevant in deriving local plasma quantities from line-integrated data, as in SXR spectroscopy, since the raw information on the plasma equilibrium, which itself is uncertain, can be combined with the raw spectroscopic data. Although outside the scope of the present work, such an approach enables self-consistent estimation of the local impurity concentrations together with the magnetic equilibrium.

b. Gaussian Process Tomography

Gaussian process tomography (GPT) is a new technique that makes use of the Bayesian framework, specifically in the choice of prior distribution $p(\bar{E}_n)$. In GPT, the prior is a Gaussian process, which imposes a level of smoothness on the emissivity field, dictated by the correlation between pixels. The choice of prior distribution is the main difference with regularization techniques based on the Fisher information.[22] Briefly, a Gaussian process (GP) is a generalization of the multivariate normal (Gaussian) distribution to a function space. It is described by a mean function $\bar{\mu}$ and a covariance function $\bar{\bar{\Sigma}}$, where $GP \sim \mathcal{N}(\bar{\mu}, \bar{\bar{\Sigma}})$. The distribution of a Gaussian process is the joint distribution of infinitely many normally distributed random variables and, as such, it is a distribution over functions over a continuous domain, e.g. time or space. GPT is related to Gaussian process regression (or "*kriging*"), a nonparametric regression technique widely used in machine learning. Being nonparametric, Gaussian

process regression does not assume any functional form for the regression function, hence leaving a lot of flexibility. Instead, the regression surface is regularized through the covariance matrix of the Gaussian process. Likewise, GPT assumes that the prior joint distribution of the emissivity in the n cells with coordinates $\bar{r}_j$ is multivariate Gaussian (Fig. 3) with covariance matrix $\bar{\bar{\Sigma}}_E$ given by:

$$\bar{\bar{\Sigma}}_E = \begin{pmatrix} k(\bar{r}_1,\bar{r}_1) & \cdots & k(\bar{r}_1,\bar{r}_n) \\ \vdots & \ddots & \vdots \\ k(\bar{r}_n,\bar{r}_1) & \cdots & k(\bar{r}_n,\bar{r}_n) \end{pmatrix}. \quad (5)$$

Here, $k(\bar{r}_j,\bar{r}_k) = cov\left[E(\bar{r}_j), E(\bar{r}_k)\right]$, with $E(\bar{r}_j) = E_j$ the emissivity in pixel j, is the *covariance kernel function*, for which we choose the common squared-exponential form:

$$k_{SE} = {\sigma_f}^2 \exp\left(-\frac{d^2}{2\sigma_l^2}\right), \quad d = \| \bar{r}_j - \bar{r}_k \|. \quad (6)$$

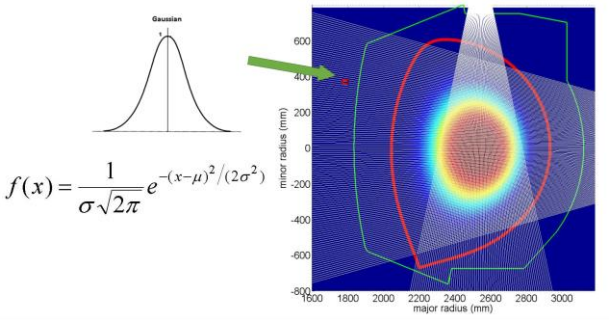

Fig. 3. In the Gaussian process framework, the emissivity in each cell follows a Gaussian distribution, while the joint distribution of every subset of pixels is multivariate normal. This imposes structure on the emissivity field, avoiding wildly fluctuating emissivity in neighboring cells. The small red square indicated in the figure represents one of the reconstruction pixels.

In turn, the kernel function depends on two parameters σ_f and σ_l, referred to as the signal standard deviation and characteristic length scale. In Bayesian terminology, the parameters of the prior distribution are called *hyperparameters* and in this case they determine the smoothness of the emissivity field. A similar role is played by the parameter governing the competition between data misfit and smoothness of the solution in minimum Fisher estimation.[19] The optimal value for that parameter is traditionally found using the L-curve method. Summarizing σ_f and σ_l by a vector $\bar{\theta}$, the total inference problem can be written as

$$p(\bar{E}_n|\bar{d}_m,\bar{\theta}) = \frac{p(\bar{d}_m|\bar{E}_n,\bar{\theta})\cdot p(\bar{E}_n|\bar{\theta})}{p(\bar{d}_m|\bar{\theta})} \sim p(\bar{d}_m|\bar{E}_n,\bar{\theta}) \cdot p(\bar{E}_n|\bar{\theta}), \quad (7)$$

where the prior is given by

$$p(\bar{E}_n|\bar{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\bar{\bar{\Sigma}}_E|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(\bar{E}_n-\bar{\mu}_E)^T\,\bar{\bar{\Sigma}}_E^{-1}\,(\bar{E}_n-\bar{\mu}_E)\right]. \quad (8)$$

Here, $\bar{\mu}_E$ is the prior mean, which will be fixed at 0, or it may be chosen on the basis of earlier experiments or expert knowledge. In principle, the hyperparameters can be marginalized from the problem (i.e. integrated out), but this would greatly increase the computational complexity of the method, thereby defeating the goal of real-time application. Instead, we will employ a common approximation wherein a fixed set of hyperparameters is determined by maximizing the evidence $p(\bar{d}_m|\bar{\theta})$, and plugging those estimates into Eq. (7). This procedure is motivated in the appendix.

The next step in the inference process consists of choosing a likelihood function $p(\bar{d}_m|\bar{E}_n,\bar{\theta})$, containing the forward model. Under the reasonable assumption of a normal distribution of the measurement uncertainty on the emissivity line integrals, described by the variable $\bar{\epsilon}$ in Eq. (2), the likelihood can be written as

$$p(\bar{d}_m|\bar{E}_n,\,\bar{\theta}) = \frac{1}{(2\pi)^{\frac{m}{2}}|\bar{\bar{\Sigma}}_d|^{\frac{1}{2}}}\, exp\left[-\frac{1}{2}(\bar{\bar{R}}\cdot\bar{E}_n-\bar{d}_m)^T\,\bar{\bar{\Sigma}}_d^{-1}\,(\bar{\bar{R}}\cdot\bar{E}_n-\bar{d}_m)\right]. \quad (9)$$

Here, $\bar{\bar{\Sigma}}_d$ is the covariance of the emissivity, describing measurement uncertainty and correlation on the vector $\bar{d}_m$ of measured line-integrals. We will assume that the various line-integrated measurements are uncorrelated and choose a 5% noise level, based on previous experience at Tore Supra. Therefore,

$$\bar{\bar{\Sigma}}_d = \begin{pmatrix} 0.05\cdot d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0.05\cdot d_m \end{pmatrix}. \quad (10)$$

Finally, the posterior distribution, conditioned on the hyperparameters, reads up to a constant factor,

$$p(\bar{E}_n|\bar{d}_m,\bar{\theta}) \sim p(\bar{d}_m|\bar{E}_n,\bar{\theta}) \cdot p(\bar{E}_n|\bar{\theta}) \sim exp\left[-\frac{1}{2}(\bar{\bar{R}}\cdot\bar{E}_n-\bar{d}_m)^T\,\bar{\bar{\Sigma}}_d^{-1}\,(\bar{\bar{R}}\cdot\bar{E}_n-\bar{d}_m)\right] exp\left[-\frac{1}{2}(\bar{E}_n-\bar{\mu}_E)^T\,\bar{\bar{\Sigma}}_E^{-1}\,(\bar{E}_n-\bar{\mu}_E)\right]. \quad (11)$$

The major advantage of normal distributions and a linear forward model now becomes clear. Indeed, it follows from standard probability calculus that the product of two normal distributions is also Gaussian, with mean vector and covariance matrix given by

$$\bar{\mu}_E^{\,post} = \bar{\mu}_E + \left(\bar{\bar{R}}^T\bar{\bar{\Sigma}}_d\bar{\bar{R}} + \bar{\bar{\Sigma}}_E^{\,-1}\right)^{-1}\bar{\bar{R}}^T\,\bar{\bar{\Sigma}}_d^{\,-1}(\bar{d}_m - \bar{\bar{R}}\cdot\bar{\mu}_E), \quad (12)$$

$$\bar{\bar{\Sigma}}_E^{post} = \left(\bar{\bar{R}}^T\bar{\bar{\Sigma}}_d\bar{\bar{R}} + \bar{\bar{\Sigma}}_E^{\,-1}\right)^{-1}. \quad (13)$$

The posterior mean is thus available in a closed form and can be used as an estimate of the emissivity field, which can be calculated in real time. In addition, the diagonal elements of the posterior covariance matrix $\bar{\bar{\Sigma}}_E^{post}$ quantify the uncertainty on the inference result. These uncertainty estimates can further guide improvements in the design of the diagnostic, e.g. by optimizing the viewing geometry (see

Section IV.b). Thus, whereas minimum Fisher information tomography relies on the maximum *a posteriori* estimate of the emissivity field, which minimizes a weighted sum of the χ^2 misfit and the Fisher information of the emissivity field, GPT yields the full posterior distribution.

It is important to note that the final reconstructed emissivity field, i.e. the posterior in Eq. (11), is relatively insensitive to the values of the hyperparameters (see the appendix). In practice, this means that, for typical plasma configurations, calculation of the hyperparameters has to be done only once. Hence, real-time SXR tomography comes within reach, since the results (12) and (13) are not based on an optimization procedure.

It should also be noted that the GPT method implemented here only uses the SXR line integral measurements and no assumptions whatsoever are made regarding the magnetic equilibrium. Likewise, no other information regarding the location of the magnetic axis or last-closed flux surface is introduced. This renders the method very flexible, potentially allowing detection of structures in the emissivity field (e.g. local impurity concentrations) that do not show up in the equilibrium reconstruction. It also prevents misguided information to enter the SXR reconstruction process in case the equilibrium information is incorrect, e.g. during sawtooth activity.

IV. GPT phantom test on WEST

Our implementation of the GPT algorithm has been validated using phantom data, i.e. a set of synthetic SXR emissivity fields characterized by various emissivity patterns, some of which pose a challenging tomography problem. Line integrals with added noise were calculated, followed by reconstruction of the emissivity field by GPT and comparison with the original emissivity phantom. In our study we employed a 5% noise level, which is the empirical reference obtained from Tore Supra measurements.

Four different shapes were used for the phantom tests, corresponding to various situations that are expected to be relevant for WEST SXR emission: the Gaussian shape, hollow shape, left-right kidney shape and up-down kidney shape, as shown in Fig. 4.

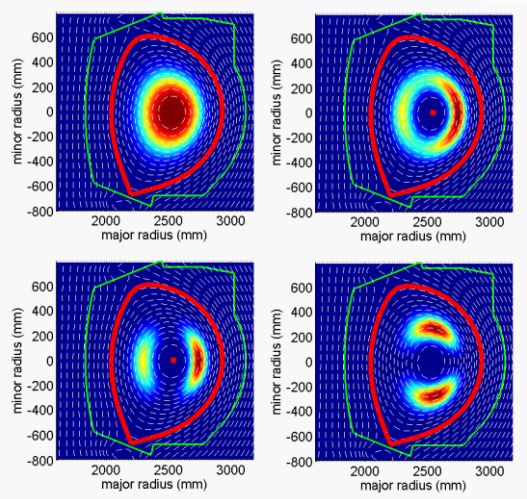

Fig. 4. Four phantom emissivity fields are used in our test: (a) Gaussian shape, (b) hollow shape, (c) left-right kidney shape, (d) up-down kidney shape. The green line in each panel represents the vacuum vessel, the white dashed lines represents the magnetic flux surfaces and the red curve is the last-closed flux surface.

a. Tomography results at 5% noise level

The reconstructed emissivity fields based on line integrals with a 5% noise level are shown in Fig. 5. The quality of the reconstructions can be quantified through a relative error map, showing the difference between the phantom and reconstructed field, normalized by the maximum phantom emissivity:

$$\xi_i = \frac{\left|E_{n,i}^{(rec)} - E_{n,i}\right|}{max\{\bar{E}_n\}}. \qquad (14)$$

In case of the Gaussian shape, the maximum relative error is around 6.9%, 15% for the hollow shape, 12% for the left-right kidney shape and 15% for the up-down kidney shape. In general, more asymmetric emissivity fields are more difficult to reconstruct, the error level depending greatly on the coverage and field of view of the optical system. Nevertheless, in all cases the characteristic shape of the phantom is recovered relatively well by GPT. In addition, one can compare the line integrals obtained from the original phantom, with those calculated from the reconstructed emissivity field. As shown in Fig. 5, good agreement is achieved in all cases. On a typical PC environment with Matlab, each time slice takes about 100 ms calculation time.

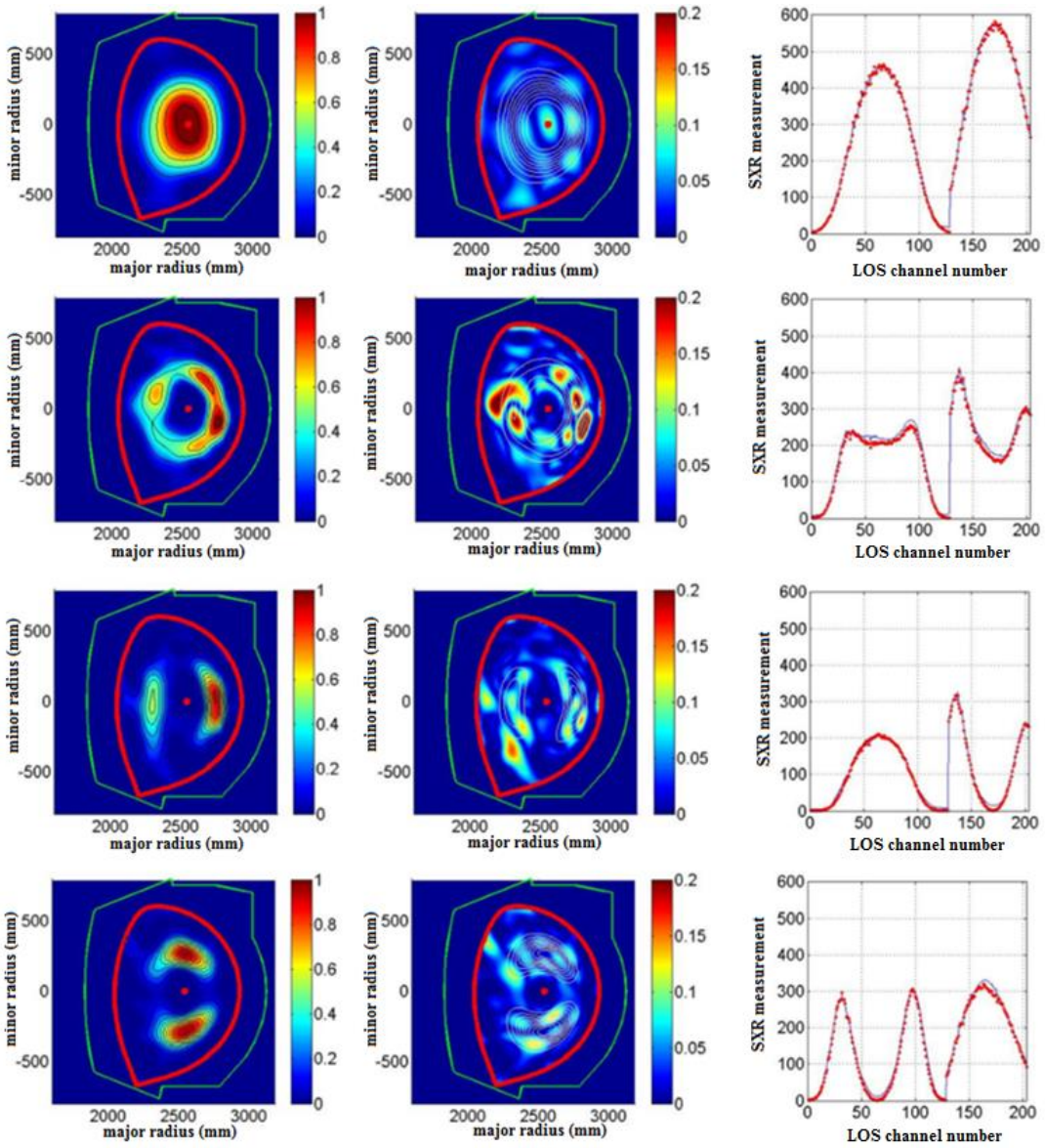

Fig. 5. GPT phantom test with 5% noise level. From left to right, the first column contains the reconstructions, the second column shows the relative error maps according to Eq. (14) (the white contours represent the original phantom), and the third column gives the comparison between the line integrals obtained from the phantom (red dots) and from the reconstructed emissivity fields (blue curves). Note that the phantom emissivity has been normalized for the benefit of numerical computation.

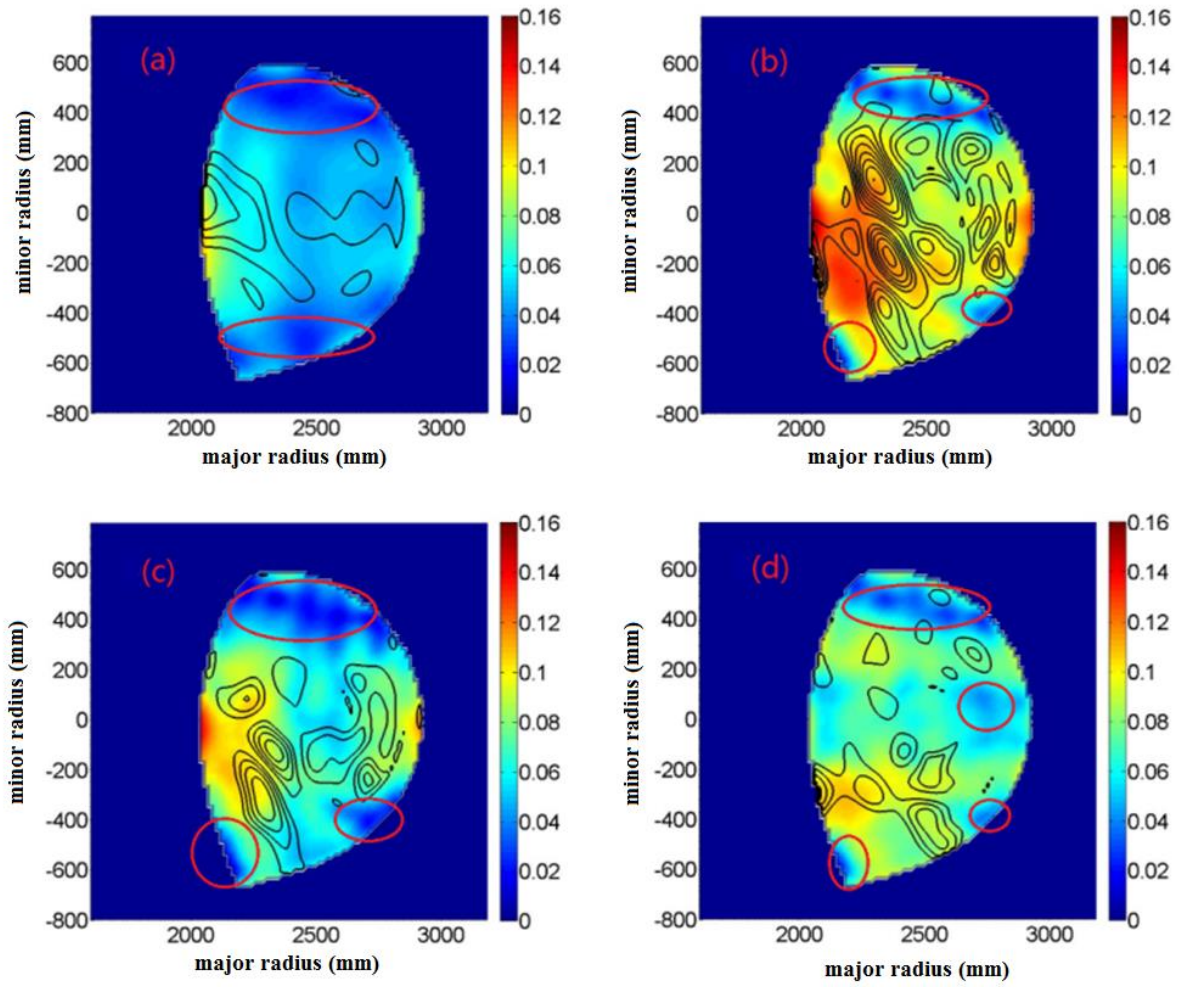

Fig. 6. Examples of a comparison between the posterior variance map (color map) and relative error map (black contours) on a 5% noise level: (a) Gaussian shape, (b) hollow shape, (c) left-right kidney shape, (d) up-down kidney shape. In the areas marked by red ellipses, both the posterior variance and relative error are low. The scale is in arbitrary units.

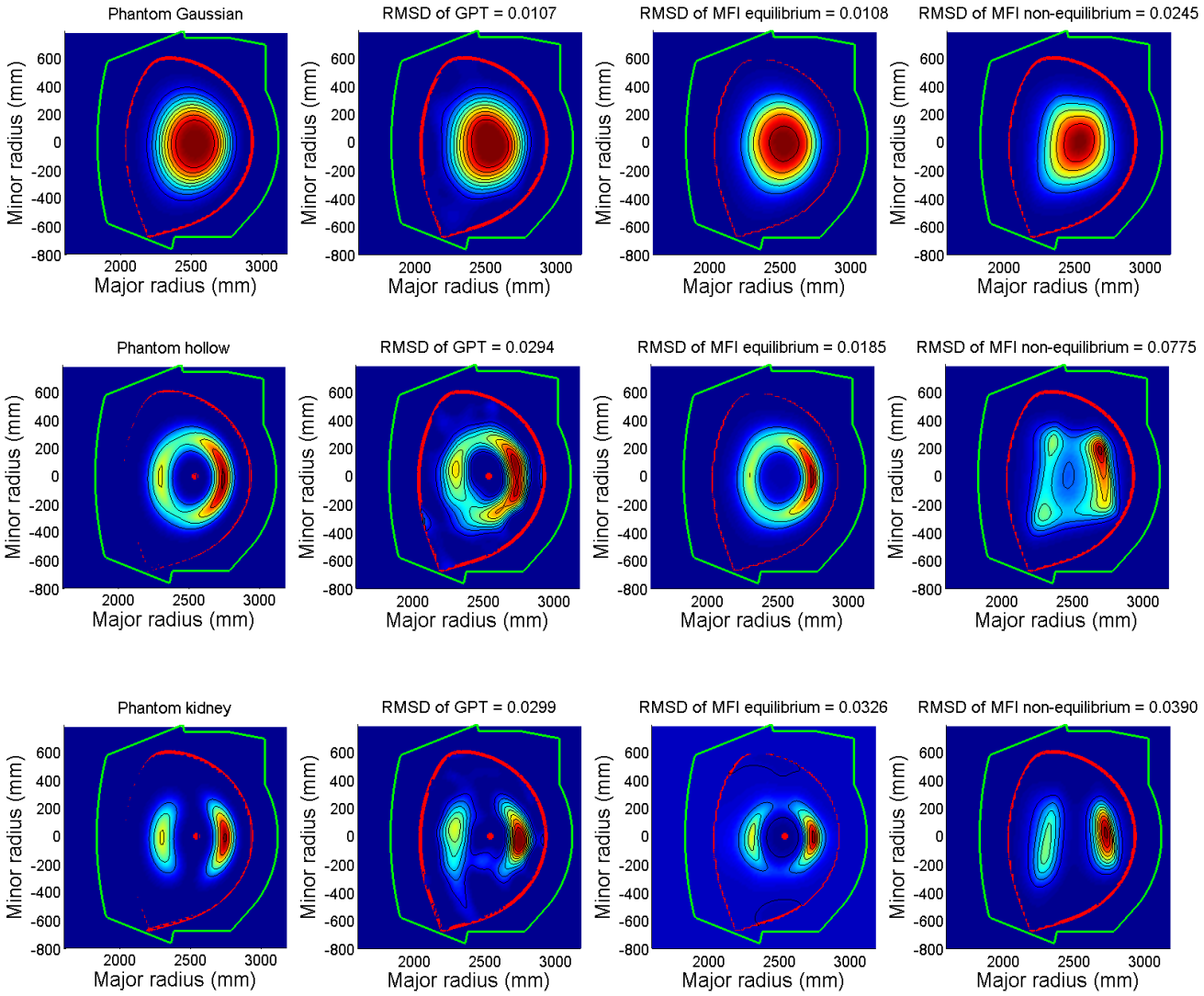

Fig. 7. Comparison of minimum Fisher information tomography and GPT. The first column shows the original phantom, the second column the results of GPT without equilibrium information, the third column minimum Fisher information tomography with equilibrium information, and the fourth column minimum Fisher information tomography without equilibrium information. Root-mean-square deviations are indicated on top of each result.

b. Gaussian process tomography uncertainty analysis

A valuable advantage of GPT is that it provides uncertainty estimates on the reconstructed emissivity field through the posterior covariance matrix; Eq. (13). This is confirmed by comparing the posterior variance map with the relative error field, as shown in Fig. 6. The uncertainty plots can be used to optimize the viewing geometry of the diagnostic, which will be part of future work. Naturally, the relative error field will not be available when performing tomography on real WEST data, but the posterior variance can still be calculated.

c. Comparison with minimum Fisher information method

Minimum Fisher information (MFI) tomography has been thoroughly tested in several fusion diagnostics, e.g. the SXR systems at JET, Tore Supra and TCV. At JET, MFI tomography has performed successfully for over 20 years. In constrast to GPT, current implementations of the minimum Fisher information technique routinely employ the magnetic equilibrium reconstruction. However, the reconstructed equilibrium is not always accurate, particularly in the core area. On the other hand, GPT in this paper does not use equilibrium information and still succeeds in producing good reconstruction results, fast enough for real time applications. Indeed, from Fig. 7 we can conclude that, even without the equilibrium assumption, GPT finds the characteristic structure of the phantoms. In our tests, MFI tomography works well when the equilibrium information is provided, but substantially worse than GPT when no such additional information is considered, relying only on the line-integrated SXR emissivity measurements. In order to quantitatively compare the quality of the reconstructions, the root-mean-square deviation (RMSD) was calculated for each result, given by

$$RMSD = \sqrt{\frac{\sum_{t=1}^{n}(E_{t,i}^{(rec)} - E_{t,i})^2}{n}}. \qquad (15)$$

These values are mentioned in Fig. 8, confirming the superiority of the GPT technique compared to minimum Fisher information in case no equilibrium information is used.

V. GPT using Tore Supra SXR data

As a validation of the GPT technique based on real SXR diagnostic data, a tomographic reconstruction of Tore Supra discharge #41864 at t = 6.2159s by GPT (no equilibrium information) and MFI (with equilibrium), is shown in Fig. 8. Since the Tore Supra SXR diagnostic system contains only 82 LOS in total, the spatial resolution was reduced to 25×25 pixels. While the reconstruction results for GPT and MFI are quite similar in the plasma core, in the boundary area GPT is less restricted by the equilibrium magnetic flux surfaces. The posterior variance and log evidence maps are shown in Fig. 9.

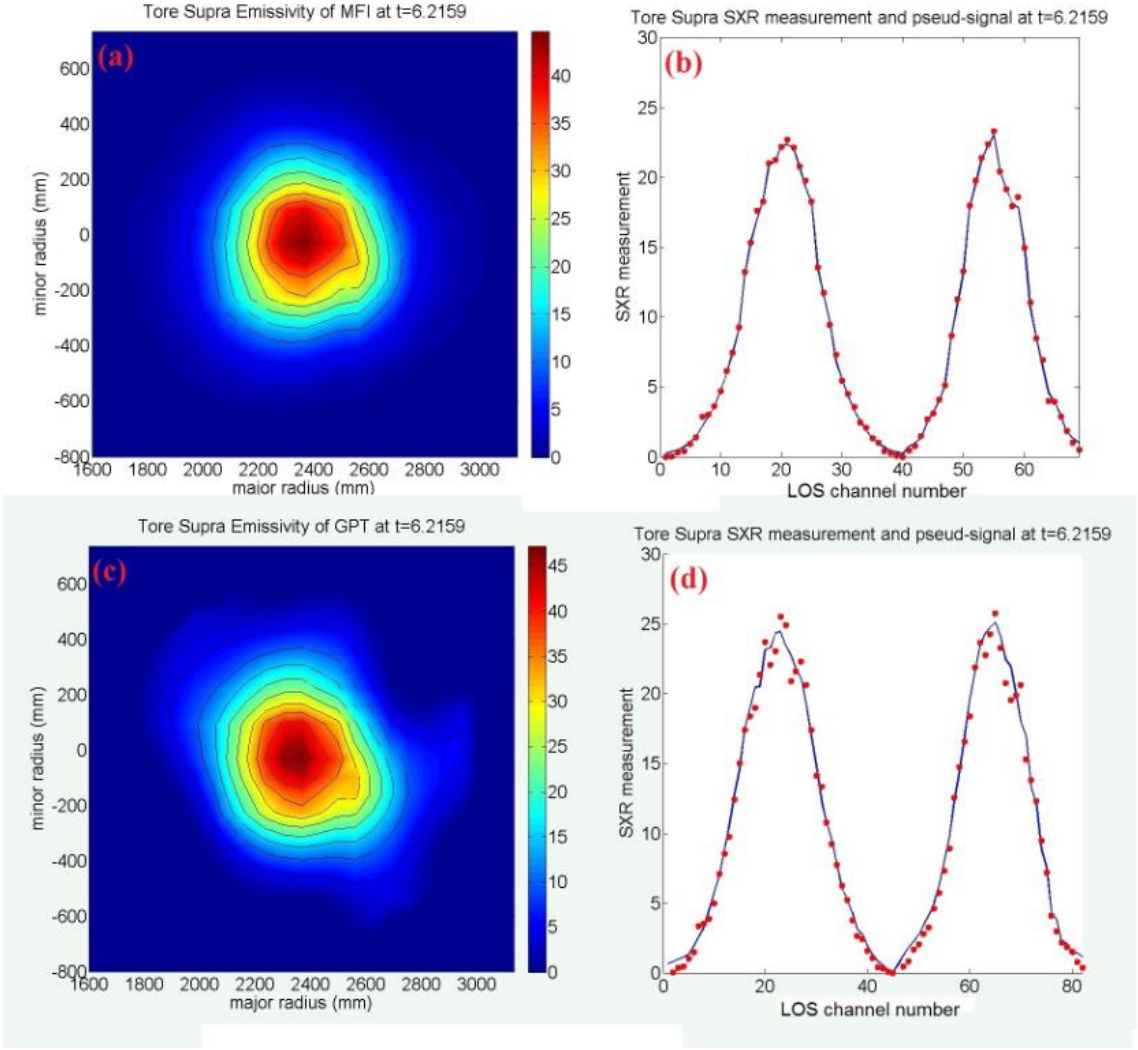

Fig. 8. Comparison of minimum Fisher information tomography and GPT based on data from Tore Supra shot #41864 at time 6.2159s. In (a) the result was obtained by GPT (no equilibrium information), while (c) shows the result of MFI with equilibrium assumption. The reconstructed emissivity fields with unit W/m^3 have been interpolated for clarity. The comparison of original and reconstructed line integrals with unit W/m^2 is also shown, where the red points are SXR measurements and the blue curves represent the reconstructed line integrals.

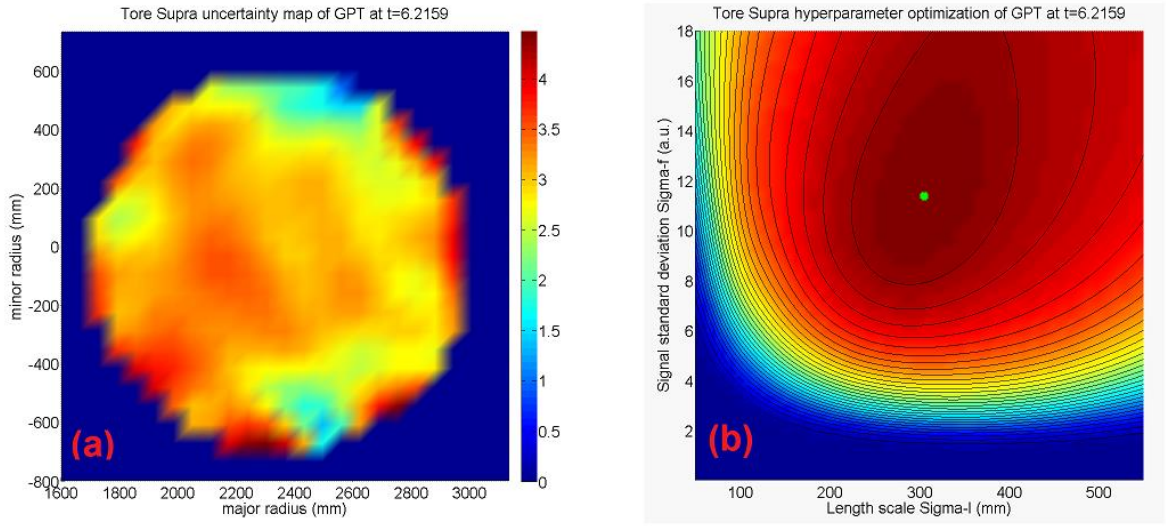

Fig. 9. (a) Posterior variance map with unit W/m^3 obtained by GPT for Tore Supra shot #41864 at time 6.2159 s. (b) Log evidence for the same data with maximum indicated by the green dot ($\sigma_f = 11.5$, $\sigma_l = 30.5$ cm).

VI. Conclusion and perspectives

In this paper, a new non-parametric SXR tomography algorithm for WEST based on Gaussian processes has been introduced. Compared to the traditional tomography techniques, GPT has several advantages. On the one hand, we have shown that the method performs well even in the absence of flux surface information from an external equilibrium reconstruction. As a result, GPT is a flexible method that can provide unbiased reconstructions. Second, GPT intrinsically provides uncertainty estimates on the reconstructed emissivity fields, obtained from the posterior Gaussian process. This can be exploited for online self-checking of the algorithm's performance. Third, the method is sufficiently fast for real-time emissivity reconstruction. Tests were carried out on four typical WEST phantom emissivity fields, yielding promising and accurate results comparing favorably to reconstructions by a standard minimum Fisher approach. The method was also shown to work well on real data from the Tore Supra SXR system, despite the reduced number of sight lines. Furthermore, the potential of the GPT method to contribute to efficient hardware design optimization was highlighted.

Although in this work we stressed the advantage of not relying on the magnetic equilibrium, this does not prevent that, in routine applications, the equilibrium may constitute a useful piece of prior information for improving the reconstruction, particularly towards the plasma boundary. Therefore, in future work we will explore the possibility to take equilibrium information into account in Gaussian process tomography, in a balanced way.

VII. Acknowledgement

This material is based upon work conducted at the Tore Supra/WEST tokamak (Institute for Magnetic Fusion Research, CEA). The work was supported by National Science Foundation of China (Nos. 11275059, 11375056, 11405049) and scholarship under the State Scholarship Fund affiliated with the Ministry of Education of the P. R. China, File No. 201508510100.

Appendix

The choice of suitable hyperparameters is a key issue for the GPT method, as they determine the degree of smoothness of the reconstructed emissivity field. A full Bayesian analysis would determine the hyperparameters together with the emissivity parameters, but this would not be feasible in real time. As a workaround, we determine the hyperparameters $\bar{\theta}$ from the data by maximizing the evidence.[1] The rationale is that the marginal posterior for the hyperparameters (i.e. with the parameters $\bar{E}_n$ marginalized) can be written as

$$p(\bar{\theta}|\bar{d}_m) \sim p(\bar{d}_m|\bar{\theta}) \cdot p(\bar{\theta}), \tag{16}$$

$\bar{\theta}$ *hyperparameters*,
$\bar{d}_m$ *GEM measurements*.

Now, assuming a non-informative uniform hyperprior distribution $p(\bar{\theta})$, we see that the posterior for the hyperparameters is proportional to the evidence $p(\bar{d}_m|\bar{\theta})$, which also occurs in Eq. (7). Hence, by maximizing the evidence w.r.t. $\bar{\theta}$, we find the maximum *a posteriori* (MAP) estimates of the hyperparameters.

To find the MAP estimates $\bar{\theta}^*$, we write the evidence as follows (see Eq. (4)):

$$p(\bar{\theta}|\bar{d}_m) \sim p(\bar{d}_m|\bar{\theta}) = \int p(\bar{d}_m|\bar{E}_n, \bar{\theta}) \;\; p(\bar{E}_n|\bar{\theta}) \;\; d\bar{E}_n \tag{18}$$

Using Eq. (11), this results in the following expression, to be maximized w.r.t. $\bar{\theta}$:

$$\log\left(p(\bar{\theta}|\bar{d}_m)\right) = -\frac{1}{2}\Big\{m\log(2\pi) + \log\|\bar{\bar{\Sigma}}_d + \bar{\bar{R}}^T\bar{\bar{\Sigma}}_E\bar{\bar{R}}\| + \bar{d}_m{}^T(\bar{\bar{\Sigma}}_d + \bar{\bar{R}}^T\bar{\bar{\Sigma}}_E\bar{\bar{R}})^{-1}\bar{d}_m\Big\}. \tag{19}$$

The hyperparameters are contained in $\bar{\bar{\Sigma}}_E$, see Eq. (5). An example of the evidence as a function of the two hyperparameters is given in Fig. 10. The data were obtained from a hollow shape phantom test and the optimization results in a length scale $\sigma_l = 14.4$ cm and signal standard deviation value $\sigma_f = 0.2427$.

Finally, once the best estimates for the hyperparameters have been found, we have to motivate plugging them into the posterior (7) or (11) for the emissivity. This posterior is obtained by marginalizing the hyperparameters from the full posterior:

$$p(\bar{E}_n|\bar{d}_m) = \int p(\bar{E}_n, \bar{\theta}|\bar{d}_m)\, d\bar{\theta} = \int p(\bar{E}_n|\bar{d}_m, \bar{\theta})\, p(\bar{\theta}|\bar{d}_m)\, d\bar{\theta}.$$

Whereas the first factor under the last integral is rather flat as a function of $\bar{\theta}$, the last factor, which we just noted to be proportional to the evidence, is usually strongly peaked. Therefore, the integral can be approximated as

$$p(\bar{E}_n|\bar{d}_m) \approx p(\bar{E}_n|\bar{d}_m, \bar{\theta}^*) \underbrace{\int p(\bar{\theta}|\bar{d}_m)d\bar{\theta}}_{=1} = p(\bar{E}_n|\bar{d}_m, \bar{\theta}^*),$$

where $\bar{\theta}^*$ refers to the MAP estimate of $\bar{\theta}$. Thus, using $\bar{\theta}^*$ in the posterior (11) is usually a justified step – commonly referred to as the *evidence approximation* in Bayesian inference.[22]

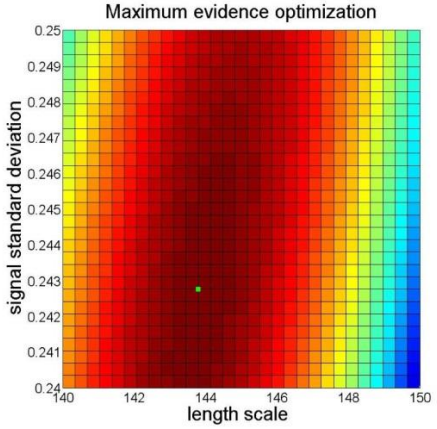

Fig. 10. An example of log evidence maximization for a 5% noise level using a hollow shape phantom test. The maximum is obtained at the location of the green dot.

References

[1] J. Svensson, *Non-parametric tomography using Gaussian processes*, JET Internal Report EFDA-JET-PR (11) 24 (2011).
[2] D. Li, J. Svensson, H. Thomsen, F. Medina, A. Werner, and R. Wolf, Rev. Sci. Instrum. **84**, 83506 (2013).
[3] T. Putterich, R. Neu, R. Dux, A.D. Whiteford, M.G. O'Mullane, and H. Summers, Nucl. Fusion **50**, 25012 (2010).
[4] D. Vezinet, D. Mazon, R. Guirlet, J. Decker, and Y. Peysson, Nucl. Fusion **54**, 83011 (2014).
[5] D. Mazon, D. Vezinet, P. Malard, D. Pacella, L. Gabelieri, A. Romano, V. Piergotti, T. Benedetto, F. Murtas, S. Dagabov, and G. Corradi, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. **720**, 78 (2013).
[6] D. Mazon, D. Vezinet, D. Pacella, D. Moreau, L. Gabelieri, A. Romano, P. Malard, J. Mlynar, R. Masset, and P. Lotte, Rev. Sci. Instrum. **83**, 063505 (2012).
[7] M.A. Dubois, D.A. Marty, and A. Pochelon, Rev. Sci. Instrum. **20**, 1355 (1980).
[8] N.R. Sauthoff, K.M. McGuire, and S. von Goeler, Rev. Sci. Instrum. **57**, 2139 (1986).
[9] P. Smeulders, Nucl. Fusion **23**, 529 (1983).
[10] Y. Nagayama, S. Tsuji, K. Kawahata, N. Noda, and S. Tanahashi, Jpn. J. Appl. Phys. **20**, L779 (1981).
[11] J.F. Camacho and R.S. Granetz, Rev. Sci. Instrum. **57**, 417 (1986).
[12] R.S. Granetz and P. Smeulders, Nucl. Fusion **28**, 457 (1988).
[13] R. Decoste, Rev. Sci. Instrum. **56**, 806 (1985).
[14] D.F. da Cruz and A.J.H. Donné, Rev. Sci. Instrum. **65**, 2295 (1994).
[15] B. Alper, S. Dillon, A.W. Edwards, R.D. Gill, R. Robins, and D.J. Wilson, Rev. Sci. Instrum. **68**, 778 (1997).
[16] M. Anton, M.J. Dutch, and H. Weisen, Rev. Sci. Instrum. **66**, 3762 (1995).
[17] L. Wang and R.S. Granetz, Rev. Sci. Instrum. **62**, 842 (1991).
[18] K. Ertl, W. von der Linden, V. Dose, and A. Weller, Nucl. Fusion **36**, 1477 (1996).
[19] M. Anton, H. Weisen, M. Dutch, P. Marmillod, and P. Paris, Plasma Phys. Control. Fusion **38**, 1849 (1996).
[20] A. Jardin, D. Mazon, M. O'Mullane, J. Mlynar, V. Loffelmann, M. Imrisek and C.B. and P.M. M. Chernyshova, T. Czarski, G. Kasprowicz, A.Wojenski, J. Instrum. **11**, C07006 (2016).
[21] D. Mazon, M. Chernyshova, G. Jiolat, T. Czarski, P. Malard, E. Kowalska-Strzeciwilk, S. Jablonski, W. Figacz, R. Zagorski, M. Kubkowska, G. Kasprowicz, K. Pozniak, W. Zabolotny, S. Larroque, J.M. Verger, M. O'Mullane, J. Mlynar, A. Byszuk, and A. Wojenski, Fusion Eng. Des. **96**–**97**, 856 (2015).
[22] W. von der Linden, V. Dose, and U. von Toussaint, *Bayesian Probability Theory: Applications in the Physical Sciences*, Cambridge University Press, 2014.